

# Structure and Content-Guided Video Synthesis with Diffusion Models

Year: 2023 | Citations: 654 | Authors: Patrick Esser, Johnathan Chiu, Parmida Atighehchian, Jonathan Granskog, Anastasis Ge

---

## Abstract

Text-guided generative diffusion models unlock powerful image creation and editing tools. Recent approaches that edit the content of footage while retaining structure require expensive re-training for every input or rely on error-prone propagation of image edits across frames. In this work, we present a structure and content-guided video diffusion model that edits videos based on descriptions of the desired output. Conflicts between user-provided content edits and structure representations occur due to insufficient disentanglement between the two aspects. As a solution, we show that training on monocular depth estimates with varying levels of detail provides control over structure and content fidelity. A novel guidance method, enabled by joint video and image training, exposes explicit control over temporal consistency. Our experiments demonstrate a wide variety of successes; fine-grained control over output characteristics, customization based on a few reference images, and a strong user preference towards results by our model.